

Weekly report (2017-10-02 ----- 2017-10-15)

Oct 15th, 2017, 21:42 pm

1. Progress

Table 1. Progress

Tasks	IN PROGRESS	Finished
Dimensionality reduction	1. Data module	Data module
	2. K-nn module (LargeVis/efanna)	K-nn module
	3. LargeVis/bh-SNE	LargeVis/bh-SNE

2. Research

2.1 paper reading

1. Distributed Representations of Words and Phrases and their Compositionality

In this paper, the authors make some extensions based on the original Skip-gram model. This method can improve both the quality of the vectors and the training speed. They also define Negative Sampling(NEG) to the hierarchical softmax and make a comparison with Noise Contrastive Estimation(NCE).

2. Distributed Representations of Sentences and Documents

In this paper, the authors propose a kind of unsupervised algorithm which is called Paragraph Vector. It can solve bag-of-words' problem-----lose orders and semantics.

3. Novel2Vec : Characterising 19th Century Fiction via Word Embeddings

In this paper, the authors mainly focus on the word embedding of supporting quantitative literary analysis (Data resources: a collection of twelve novels from three 19th century novelists - six by Jane Austen, three by Charles Dickens, and three by Arthur Conan Doyle). The authors analyze the effects of applying two variants of word2vec, a continuous-bag-of-words strategy and a skip-gram strategy.